



## MODELING SURVIVAL DATA USING COX PROPORTIONAL HAZARD MODELS

Senthilkumar V<sup>1</sup> and Sachithanatham S<sup>2</sup><sup>1</sup>Manonmaniam Sundaranar University, Tirunelveli – 627 012<sup>2</sup>Department of Statistics, Arignar Anna Government Arts College, Villupuram – 605602

## ARTICLE INFO

**Article History:**Received 13<sup>th</sup> January, 2018Received in revised form 12<sup>th</sup>February, 2018 Accepted 5<sup>th</sup> March, 2018Published online 28<sup>th</sup> April, 2018

## ABSTRACT

It was pointed out that in the literature Cox (1972) proposed a model and it is not based on conditional likelihood and there may be same flaws in logic. Cox (1975) was able to recast this method of estimation through “partial likelihood”. This approach seemed to be based on sound inferential principles. Following the approach of Cox (1975) it is proposed to study the proportional hazard model for survival analysis. Numerical results are substantiated through real data example.

**Copyright © Senthilkumar V and Sachithanatham S 2018**, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

It was pointed out that in the literature Cox (1972) proposed a model and it is not based on conditional likelihood and there may be same flaws in logic. Cox (1975) was able to recast this method of estimation through “partial likelihood”. This approach seemed to be based on sound inferential principles. Following the approach of Cox (1975) it is proposed to study the proportional hazard model for survival analysis. Numerical results are substantiated through real data example

**Multiple Myeloma**

Multiple myeloma is form of blood cancer that develops in the bone marrow. In multiple myeloma, normal plasma cells transform into malignant myeloma cells and produce large quantities of an abnormal immunoglobulin called monoclonal protein or M protein. The monoclonal protein produced by myeloma cells, interferes with normal blood cell production. In addition, the levels of functional immunoglobulins are depressed in individuals with multiple myeloma. Although the process is not completely understood, it appears that the functional immunoglobulins made by existing, healthy plasma cells breaks down more quickly in patients with multiple myeloma than in healthy individuals. Normal plasma cells (B cells) transform into malignant plasma cells (myeloma cells) through a multistep process.

The process begins when specific adhesion molecules on the surface of myeloma cells allow the cells to bind to stromal cells in the bone marrow. This attachment allows both the stromal cells and the myeloma cells to produce chemical messengers called cytokines that cause the further development of myeloma.

1. Stromal cells produce an increased amount of the cytokine interleukin-6 (IL-6), which is necessary for the continued growth and survival of the myeloma cells. The high level of IL-6 leads to an uncontrolled increase in the number of myeloma cells.
2. Myeloma cells produce an increased amount of IL-1-beta and tumor necrosis factor-beta (TNF-beta), two cytokines that activate the development of osteoclasts, cells that break down bone as part of normal bone remodeling. However, in multiple myeloma, the increased number of osteoclasts causes damage to bone.
3. Myeloma cells also produce growth factors, such as vascular endothelial growth factor (VEGF), that promote angiogenesis, or the creation of new blood vessels. These new blood vessels bring oxygen and nutrients to the tumor, helping it to grow by generating new myeloma cells.

**The Proportional Hazards Model**

A proportional hazards model proposed by Cox (1972) assumes that

$$\lambda\left(\frac{t}{z}\right) = \lambda_0(t)e^{z_1\beta_1+\dots+z_p\beta_p} = \lambda_0(t)e^{z^T\beta} \quad \dots (1)$$

\*Corresponding author: Senthilkumar V

Manonmaniam Sundaranar University, Tirunelveli – 627 012

where  $z$  is a  $p \times 1$  vector of covariates such as treatment indicators, prognostic factors, etc., and  $\beta$  is a  $p \times 1$  vector of regression coefficients. Note that there is no intercept  $\beta_0$  in model in equ. (1).

Obviously,

$$\lambda\left(\frac{t}{z} = 0\right) = \lambda_0(t)$$

So  $\lambda_0(t)$  is often called the baseline hazard function. It can be interpreted as the hazard function for the population of subjects with  $z = 0$ . The baseline hazard function  $\lambda_0(t)$  in model in equ. (1) can take any shape as a function of  $t$ . The only requirement is that  $\lambda_0(t) > 0$ . This is the nonparametric part of the model and  $z^T \beta$  is the parametric part of the model. So Cox's proportional hazards model is a semiparametric model.

### Interpretation of a Proportional Hazards Model

It is easy to show that under model in equ. (1)

$$S\left(\frac{t}{z}\right) = [S_0(t)]^{\exp(z^T \beta)}$$

where  $S(t/z)$  is the survival function of the subpopulation with covariate  $z$  and  $S_0(t)$  is the survival function of baseline population ( $z = 0$ ). That is

$$S_0(t) = e^{-\int_0^t \lambda_0(u) du}$$

For any two sets of covariates  $Z_0$  and  $Z_1$ ,

$$\frac{\lambda\left(\frac{t}{z_1}\right)}{\lambda\left(\frac{t}{z_0}\right)} = \frac{\lambda_0(t)e^{z_1^T \beta}}{\lambda_0(t)e^{z_0^T \beta}} = e^{(z_1 - z_0)^T \beta} \text{ for all } t \geq 0,$$

which is a constant over time (so the name of proportional hazards model). Equivalently,

$$\log \left[ \frac{\lambda\left(\frac{t}{z_1}\right)}{\lambda\left(\frac{t}{z_0}\right)} \right] = (z_1 - z_0)^T \beta \text{ for all } t \geq 0,$$

With one unit increase in  $z_k$  while other covariate values being held fixed, then

$$\log \left[ \frac{\lambda\left(\frac{t}{z_k + 1}\right)}{\lambda\left(\frac{t}{z_k}\right)} \right] = \log \left( \lambda\left(\frac{t}{z_k + 1}\right) \right) - \log \left( \lambda\left(\frac{t}{z_k}\right) \right) = \beta_k$$

Therefore,  $\beta_k$  is the increase in log hazard (i.e., log hazard-ratio) at any time with unit increase in the  $k^{\text{th}}$  covariate  $z_k$ . Equivalently,

$$\frac{\lambda\left(\frac{t}{z_k + 1}\right)}{\lambda\left(\frac{t}{z_k}\right)} = e^{\beta_k}, \text{ for all } t \geq 0$$

So  $\exp(\beta_k)$  is the hazard ratio associated with one unit increase in  $Z_k$ . Furthermore, since

$$P[t \leq T < t + \Delta t | T \geq t, z] \approx \lambda\left(\frac{t}{z}\right) \Delta t, \text{ we have}$$

$$\frac{P[t \leq T < t + \Delta t | T \geq t, z_k + 1]}{P[t \leq T < t + \Delta t | T \geq t, z_k]} \approx e^{\beta_k}, \text{ for all } t \geq 0$$

so  $\exp(\beta_k)$  can be loosely interpreted as the ratio of two conditional probabilities of dying in the near future given a subject is alive at any time  $t$ . Since

$$\frac{\lambda\left(\frac{t}{z_k + 1}\right) - \lambda\left(\frac{t}{z_k}\right)}{\lambda\left(\frac{t}{z_k}\right)} = e^{\beta_k} - 1$$

So  $e^{\beta_k} - 1$  can be interpreted as the percentage change (increase or decrease) in hazard with one unit increase in  $z_k$  while adjusting for other covariates.

### Estimation using Partial Likelihood

Data and Model

1. Data:  $(X_i, \Delta_i, Z_i), i = 1, \dots, n$ , where for the  $i^{\text{th}}$  individual

$$X_i = \min(T_i, C_i).$$

$$\Delta_i = I(T_i \leq C_i),$$

$$Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{ip})^T \text{ is a vector of covariates.}$$

2. Proportional hazards model

$$\lambda\left(\frac{t}{z_i}\right) = \lambda_0(t)e^{z_i^T \beta}$$

where

$$\lambda\left(\frac{t}{z_i}\right) = \lim_{h \rightarrow 0^+} \left\{ \frac{P[t \leq T_i < t + h | T_i \geq t, z_i]}{h} \right\}$$

Assume that  $C_i$  and  $T_i$  are conditionally independent given  $Z_i$ . Then the cause-specific hazard can be used to represent the hazard of interest. That is (in terms of conditional probabilities)

$$\begin{aligned} P[x \leq X_i < x + \Delta x, \Delta_i = 1 | X_i \geq x, z_i] \\ = P[x \leq T_i < x + \Delta x | T_i \geq x, z_i] \\ \approx \lambda_{T_i}(x | z_i) \Delta x \end{aligned}$$

Similar to the case of log rank test. Assume the survival time is continuous. Hence can take the grid points dense enough so that at most one death can occur within any interval.

### Notations

1.  $dN_i(u)$  denote the indicator for the  $i^{\text{th}}$  individual being observed to die in  $[u, u + \Delta u)$ .  
 $dN_i(u) = I(X_i \in [u, u + \Delta u), \Delta_i = 1)$ .
2.  $Y_i(u)$  denote the indicator for whether or not the  $i^{\text{th}}$  individual is at risk at time  $u$ .
3.  $Y_i(u) = I(X_i \geq u)$   $dN(u) = \sum_{i=1}^n dN_i(u)$  denote the number of deaths for the whole sample occurring in  $[u, u + \Delta u)$ . Since we are assuming  $\Delta u$  is sufficiently small, so  $dN(u)$  is either 1 or 0 at any time  $u$ .
4.  $Y(u) = \sum_{i=1}^n Y_i(u)$  be the total number from the entire sample who are at risk at time  $u$ .
5.  $F(x)$  denote the information up to time  $x$  (one of the grid points)  $F(x) = \{(dN_i(u), Y_i(u), z_i), i = 1, \dots, n \text{ for grid points } u < x \text{ and } dN(x)\}$  is the individual who was observed to die among those at risk at time  $x$  if  $dN(x) = 1$ .
6.  $I(x)$  denote the individual in the sample who died at time  $x$  if someone died. If no one dies at time  $x$ , then  $I(x) = 0$  and if  $I(x) = j$ , then this means

that the  $j^{\text{th}}$  individual in the sample with covariate vector  $Z_j$  died in  $[x, x+\Delta x)$ .

7.  $F(\infty)$  denote all the data in the sample.  $F(\infty) = \{(X_i, \Delta_i, z_i), i = 1, \dots, n\}$
8. If we let  $u_1 < u_2 < \dots$  denote the value of the grid points along the time axis, then the data (with redundancy) can be expressed as  $(F(u_1), I(u_1), (u_2), I(u_2), \dots, F(\infty))$ .

**RESULT**

Denote the observed values of the above random variables by lower cases. Then the likelihood of the parameter  $\lambda_0(t)$  and  $\beta$  can be written as

$$P[F(u_1) = f(u_1); \lambda_0(\cdot), \beta] \times P[I(u_1) = i(u_1) | F(u_1) = f(u_1); \lambda_0(\cdot), \beta] \times P[F(u_2) = f(u_2) | F(u_1) = f(u_1), I(u_1) = i(u_1); \lambda_0(\cdot), \beta] \times P[I(u_2) = i(u_2) | F(u_1) = f(u_1), I(u_1) = i(u_1), F(u_2) = f(u_2); \lambda_0(\cdot), \beta] \times \dots$$

and the last term can be simplified as

$$P[I(u_2) = i(u_2) | F(u_1) = f(u_1), I(u_1) = i(u_1), F(u_2) = f(u_2); \lambda_0(\cdot), \beta] = P[I(u_2) = i(u_2) | F(u_2) = f(u_2); \lambda_0(\cdot), \beta]$$

That is, the full likelihood can be written as the product of a series of conditional likelihoods.

Obviously, the log partial likelihood function of  $\beta$  is

$$l(\beta) = \sum_{\{all\ grid\ pts\ u\}} dN(u) \left[ z_{I(u)} \beta - \log \left( \sum_{l=1}^n \exp(z_l \beta) Y_l(u) \right) \right]$$

The score function is

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{\{all\ grid\ pts\ u\}} dN(u) \left[ z_{I(u)} - \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} \right]$$

and the second derivative is

$$\frac{\partial^2 l(\beta)}{\partial \beta^2} = - \sum_u dN(u) \left[ \frac{\sum_{l=1}^n z_l^2 \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} - \left( \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} \right)^2 \right]$$

Define

$$\bar{z}(u, \beta) = \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} = \sum_{l=1}^n z_l w_l$$

where

$$w_l = \frac{\exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)}$$

is the weight that is proportional to the hazard of the individual failing. So  $\bar{z}(u, \beta)$  can be interpreted as the weighted average of the covariate  $z$  among those individuals still at risk at time  $u$  with weights  $w_l$  and

$$V_z(u, \beta) = \left[ \frac{\sum_{l=1}^n z_l^2 \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} - \left( \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} \right)^2 \right]$$

$$= \left[ \frac{\sum_{l=1}^n z_l^2 \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} - (\bar{z}(u, \beta))^2 \right] = \sum_{l=1}^n z_l^2 w_l - (\bar{z}(u, \beta))^2$$

This can be shown to be equal to

$$V_z(u, \beta) = \sum_{l=1}^n \left[ \frac{(z_l - \bar{z}(u, \beta))^2 \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} \right] = \sum_{l=1}^n z_l - (\bar{z}(u, \beta))^2 w_l$$

This last representation says that  $V_z(u, \beta)$  can be interpreted as the weighted variance of the covariates among those individuals still at risk at  $u$  and hence  $V_z(u, \beta) > 0$ . Consequently,

$$\frac{\partial^2 l(\beta)}{\partial \beta^2} = - \sum_u dN(u) V_z(u, \beta) < 0.$$

The above results can be displayed graphically in Fig. 1.

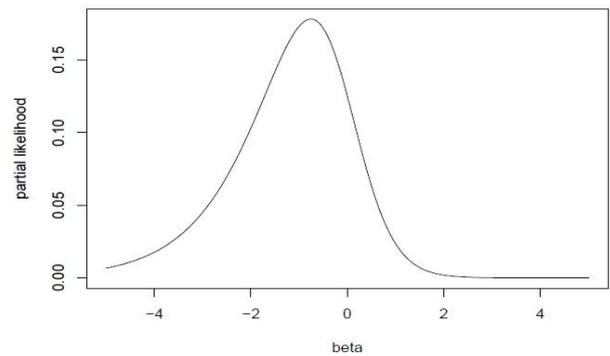


Fig 1 The partial likelihood

Therefore  $l(\beta)$  has a unique maximizer and can be obtained uniquely by solving the following partial likelihood equation:

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{\{all\ grid\ pts\ u\}} dN(u) \left[ z_{I(u)} - \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} \right] = 0$$

This maximizer  $\hat{\beta}$  defines the MPLE of  $\beta$

Terminology: The quantity

$$- \frac{\partial^2 l(\beta)}{\partial \beta^2} = \sum_u dN(u) V_z(u, \beta)$$

is defined as the partial likelihood observed information and is denoted by  $J(\beta)$ . and MPLE  $\hat{\beta}$  has the following statistical properties has to be satisfied for getting the result.

1. Consistency: That is,  $\hat{\beta}$  will converge to the true value of  $\beta$  which generated the data as the sample size gets larger. We call this true value  $\beta_0$
2. Asymptotic Normality:  $\hat{\beta}$  will be approximately normally distributed with mean  $\beta_0$  and a variance which can be estimated from the data. This approximation will be better as the sample size gets larger. This result is useful in making inference for the true  $\beta$ .
3. Efficiency: Among all other competing estimators for  $\beta$ , the MPLE has the smallest variance, at least, when the sample size gets larger.

In order to show the properties for  $\hat{\beta}$ , expand  $U(\hat{\beta})$  at the true value  $\beta_0$  using Taylor expansion:

$$0 = U(\hat{\beta}) \approx U(\beta_0) + \frac{\partial U(\beta_0)}{\partial \beta} (\hat{\beta} - \beta_0)$$

Since

$$\frac{\partial U(\beta_0)}{\partial \beta} = \frac{\partial^2 l(\beta_0)}{\partial \beta^2} = -J(\beta_0)$$

therefore

$$(\hat{\beta} - \beta_0) \approx [J(\beta_0)]^{-1} U(\beta_0)$$

This expression indicates that we need to investigate the properties of the score function  $U(\beta_0)$

$$U(\beta_0) = \sum_u dN(u) [z_{l(u)} - \bar{z}(u, \beta_0)]$$

Further it is evident that

$$E[U(\beta_0)] = 0$$

Since

$$E[U(\beta_0)] = E[\sum_u dN(u) (z_{l(u)} - \bar{z}(u, \beta_0))] \quad \text{and}$$

$$E[dN(u) (z_{l(u)} - \bar{z}(u, \beta_0))] = E[E[dN(u) (z_{l(u)} - \bar{z}(u, \beta_0)) | F(u)]]$$

Conditional on  $F(u)$ ,  $dN(u)$  and  $\bar{z}(u, \beta_0)$  are both known.

Consequently the inner expectation can be written as

$$dN(u) [E[z_{l(u)} | F(u)] - \bar{z}(u, \beta_0)]$$

Remember that  $I(u)$  is the patient identifier for the individual that dies at time  $u$  and is set to zero if no one dies at  $u$ . If no one dies at  $u$ , then  $dN(u) = 0$ , and hence the above quantity is zero. If someone dies at  $u$ , then  $dN(u) = 1$ , and conditional on  $F(u)$ , we know it has to be one of the  $Y(u)$  people at risk at time  $u$ ; i.e.,  $I(u)$  must be one of the values  $\{i: Y_i = 1\}$ .

The conditional distribution of  $Z_l(u)$  given  $F(u)$  can be derived through the conditional distribution of  $I(u)$  given  $F(u)$  as shown in Table 1.

Therefore

$$E[z_{l(u)} | F(u)] = \sum_{l=1}^n z_l w_l = \frac{\sum_{l=1}^n z_l \exp(z_l \beta) Y_l(u)}{\sum_{l=1}^n \exp(z_l \beta) Y_l(u)} = \bar{z}(u, \beta_0)$$

From this, we immediately get

$$E[U(\beta_0)] = 0$$

**Table 1** Conditional distribution of  $zI(u)$  given  $F(u)$

Values of $I(u)$	Values of $z_l(u)$	Probability
1	$z_1$	$\frac{\exp(z_1 \beta_0) Y_1(u)}{\sum_{l=1}^n \exp(z_l \beta_0) Y_l(u)} = w_1$
2	$z_2$	
.	.	
.	.	
n	$z_n$	$\frac{\exp(z_n \beta_0) Y_n(u)}{\sum_{l=1}^n \exp(z_l \beta_0) Y_l(u)} = w_n$

The unbiased estimate for the variance of  $U(\beta_0)$

$E[U(\beta_0)] = 0$ , so

$$\text{Var}[U(\beta_0)] = E[U(\beta_0)]^2$$

$$= E \left[ \sum_u dN(u) (z_{l(u)} - \bar{z}(u, \beta_0)) \right]^2$$

$$= E \left[ \sum_u \{dN(u) (z_{l(u)} - \bar{z}(u, \beta_0))\}^2 \right] + E \left[ \sum_{u \neq u'} \{dN(u) (z_{l(u)} - \bar{z}(u, \beta_0))\} \{dN(u') (z_{l(u')} - \bar{z}(u', \beta_0))\} \right]$$

Assume  $u' > u$  and denote

$$A(u) = dN(u) [z_{l(u)} - \bar{z}(u, \beta_0)], \quad A(u') = dN(u') [z_{l(u')} - \bar{z}(u', \beta_0)]$$

Then the expectation of the cross-product is

$$E[A(u)A(u')] = E[E[A(u)A(u')/F(u')]]$$

Since  $u' > u$ , conditional on  $F(u')$ ,  $A(u)$  is known. So

$$E[A(u)A(u')/F(u')] = A(u)E[A(u')/F(u')] = 0$$

Therefore

$$\text{Var}[U(\beta_0)] = E \sum_u [A^2(u)]$$

$$= \sum_u E[A^2(u)]$$

$$= \sum_u E[E[A^2(u) | F(u)]]$$

The inner conditional expectation is

$$E[A^2(u) | F(u)] = E[\{dN(u) (z_{l(u)} - \bar{z}(u, \beta_0))\}^2 | F(u)]$$

$dN(u)$  is either 0 or 1, so  $dN^2(u) = dN(u)$ . Hence

$$E[A^2(u) | F(u)] = E[dN(u) \{z_{l(u)} - \bar{z}(u, \beta_0)\}^2 | F(u)]$$

Conditional on  $F(u)$ ,  $dN(u)$  is known,  $\bar{z}(u, \beta_0)$  is also known and from Table 1

$$\bar{z}(u, \beta_0) = E[z_{l(u)} | F(u)]$$

Therefore

$$E[A^2(u) | F(u)] = dN(u) E[\{z_{l(u)} - \bar{z}(u, \beta_0)\}^2 | F(u)] = dN(u) \text{var}[z_{l(u)} | F(u)] = dN(u) V_z(u, \beta_0)$$

Consequently,

$$\text{Var}[U(\beta_0)] = \sum_u E[dN(u) V_z(u, \beta_0)]$$

$$= E[\sum_u dN(u) V_z(u, \beta_0)]$$

Note that the quantity  $\sum_u dN(u) V_z(u, \beta_0)$  is a statistic (can be calculated from the observed data), so  $\sum_u dN(u) V_z(u, \beta_0)$  is an unbiased estimate of  $\text{Var}[U(\beta_0)]$ . In fact,  $\sum_u dN(u) V_z(u, \beta_0)$  is the partial likelihood observed information  $J(\beta_0)$ .

### Numerical Illustration

#### Inference with a Single Covariate

Assume a proportional hazards model with a single covariate  $z$

$$\lambda(t) = \lambda_0(t) e^{z\beta}$$

After we get our data  $(x_i, \delta_i, z_i)$ , we can obtain the MPLE  $\hat{\beta}$  by solving the partial likelihood equation; i.e., setting the partial score to zero. Then asymptotically,

$$\hat{\beta} \sim N(\beta_0, J^{-1}(\hat{\beta}))$$

We can use this fact to construct confidence interval for  $\beta$  and test the hypothesis  $H_0: \beta = \beta_0$ ,

etc. For example, a  $(1 - \alpha)$  CI of  $\beta$  is

$$\hat{\beta} \pm z_{\alpha/2} [J^{-1}(\hat{\beta})]^{1/2}$$

myelomatosis data with treatments 1 and 2. Define a treatment indicator  $\text{trt1}$  which takes value 0 for treatment

1 and takes value 1 for treatment 2. Using SAS Program, the data were analysis

**Comparison of score test and two-sample log rank test**

Assume  $z$  is the dichotomous indicator for treatment; i.e.,

$$Z = \begin{cases} 1 & \text{for treatment 1} \\ 0 & \text{for treatment 0} \end{cases}$$

and the proportional hazards model:

$$\lambda(t) = \lambda_0(t)e^{z\beta}$$

Score test: Under  $H_0: \beta = 0$ , the score  $U(0)$  (evaluated under  $H_0$ ) has the distribution

$$U(0) \sim N(0, J(0)).$$

Or equivalently,

$$\left[ \frac{U(0)}{\sqrt{J(0)}} \right]^2 \sim \chi^2_1$$

Since the score  $U(0)$  has the expression

$$U(0) = \sum_u dN(u) [z_{1(u)} - \bar{z}(u, 0)]$$

Then

1. If a death occurs at time  $u$ , then  $dN(u) = 1$ , in which case there will a contribution to  $U(0)$  by adding  $[z_{1(u)} - \bar{z}(u, 0)]$ . Otherwise no contribution.
2. Since  $z = 1$  for treatment 1 and  $z = 0$  for treatment 0,  $z_1(u)$  will then the number of deaths at time  $u$  from treatment 1.
3. Under  $H_0: \beta = 0$ ,  $\bar{z}(u, 0)$  is simplified to be

$$\bar{z}(u, 0) = \frac{\sum_{l=1}^n z_l Y_l(u)}{\sum_{l=1}^n Y_l(u)}$$

which is the proportion of individuals in group 1 among those at risk at time  $u$ . Since we only assume one death at time  $u$ , this proportion is the expected number of death for treatment 1 among those at risk at time  $u$ , under the null hypothesis of no treatment difference.

4. Therefore,  $U(0)$  is the sum over the death times of the observed number of deaths from treatment 1 minus the expected number of deaths under the null hypothesis. This was the numerator of the two-sample log rank test:

$$\sum_u \left[ dN_1(u) - \frac{Y_1(u)}{Y(u)} dN(u) \right]$$

where  $dN_1(u) = \#$  of observed deaths from treatment 1,  $Y_1(u) = \#$  at risk at time  $u$  from treatment 1,  $Y(u) =$  total # at risk at time  $u$  from 2 treatments,  $dN(u) =$  total # of deaths from 2 treatments.

5. The denominator of the score test was computed as

$$J^{1/2}(0) = \left[ \sum_u dN(u) V_z(u, 0) \right]^{1/2}$$

where

$$V_z(u, 0) = \frac{\sum_l (z_l - \bar{z}(u, 0))^2 Y_l(u)}{\sum_l Y_l(u)}$$

Among the  $Y(u)$  individuals at risk at time  $u$ , there are  $Y_1(u)$  individuals whose  $Z_l$  value of  $Z_l = 1$  and  $Y_0(u)$  individuals whose  $Z_l$  value of  $Z_l = 0$ . We already argued that

$$\bar{z}(u, 0) = \frac{Y_1(u)}{Y(u)}$$

Therefore,

$$V_z(u, 0) = \frac{\sum_l (z_l - \bar{z}(u, 0))^2 Y_l(u)}{\sum_l Y_l(u)} = \frac{\left[ 1 - \frac{Y_1(u)}{Y(u)} \right]^2 Y_1(u) + \left[ 0 - \frac{Y_1(u)}{Y(u)} \right]^2 Y_0(u)}{Y(u)} \quad (z_1(u) \text{ takes 1 or 0})$$

$$= \frac{\frac{Y_0^2(u) Y_1(u)}{Y^2(u)} + \frac{Y_1^2(u) Y_0(u)}{Y^2(u)}}{Y(u)} (Y_1(u) + Y_0(u) = Y(u)) = \frac{Y_1(u) Y_0(u) Y(u)}{Y^3(u)} = \frac{Y_1(u) Y_0(u)}{Y^2(u)}$$

Therefore

$$J(0) = \sum_u dN(u) - \frac{Y_0(u) Y_1(u)}{Y^2(u)}$$

The variance is used to compute the logrank test statistic:

$$\sum_u \left[ \frac{Y_1(u) Y_0(u) dN(u) [Y(u) - dN(u)]}{Y^2(u) [Y(u) - 1]} \right]$$

In the special case where  $dN(u)$  can only be one or zero, then above expression reduces to

$$\sum_u \left[ \frac{Y_1(u) Y_0(u) dN(u) [Y(u) - 1]}{Y^2(u) [Y(u) - 1]} \right] = \sum_u \left[ \frac{Y_1(u) Y_0(u) dN(u)}{Y^2(u)} \right]$$

which is exactly equal to  $J(0)$ .

Therefore, continuous survival time data with no ties, the score test of the hypothesis  $H_0: \beta = 0$  in the proportional hazards model is exactly the same as the logrank test for dichotomous covariate  $z$ .

The score test

$$\left[ \frac{U(0)}{\sqrt{J(0)}} \right]^2$$

can be used to test the hypothesis  $H_0: \beta = 0$  for the model  $\lambda(t|z) = \lambda_0(t)e^{z\beta}$  for any covariate value  $z$ , whether or not  $z$  is discrete or continuous. The null hypothesis  $H_0: \beta = 0$  implies that the hazard rate at any time  $t$  is unaffected by the covariate  $z$ . This also implies that the survival distribution does not depend on  $z$ . The alternative hypothesis  $H_A: \beta \neq 0$  implies that hazard rate increases or decreases (depending on the sign of  $\beta$ ) as  $z$  increases throughout all time. Therefore, belief in this alternative hypothesis would mean that individuals with a higher value of  $z$  would have stochastically larger (or smaller depending on the sign of  $\beta$ ) survival distribution than those individuals with a smaller values of  $z$ . The test command in Proc Lifetest computes the score test of the hypothesis  $H_0: \beta = 0$  for the proportional hazards model. Consequently, when using the test command, the covariate  $z$  is not limited to being dichotomous, nor discrete.

The treatment difference between treatments 1 and 2 for myelomatosis data given in Krall et. al. (1975) and Kailash (1992) has been used. Using SAS and the following SAS commands used and results were shown in table 1, table 2, and table 3.

```
proc lifetest data=myel;
time dur*status(0);
```

test trt;  
run;  
and part of the output is presented in the following:  
Univariate Chi-Squares for the LOG RANK Test

**Table 1** Chi-square for logRank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	p
TRT	-1.8737	1.587	0.9750	0.1471

**Table 2** Covariance Matrix for the LOG RANK Statistics

Variable	TRT
TRT	4.21151

**Table 3** Stepwise Sequence of Chi-Squares for the LOG RANK Test

Variable	DF	Chi-Square	p	Chi-Square Increment	p
TRT	1	1.0953	0.1476	1.0759	0.1475

**Likelihood Ratio Test**

As in the ordinary likelihood theory, the (partial) likelihood ratio test can also be used to test the null hypothesis:

$$H_0: \beta = \beta_0.$$

Recall that  $l(\beta)$  is the log partial likelihood. Intuitively, if  $H_0$  is true, then  $\hat{\beta}$ , the MPLE of  $\beta$ , should be close to  $\beta_0$ . Hence  $l(\hat{\beta})$  should be close to  $l(\beta_0)$ . Since  $l(\hat{\beta}) - l(\beta_0)$  is always non-negative, so we should reject  $H_0$  when this difference is large.

The likelihood ratio test uses the fact that  $2[l(\hat{\beta}) - l(\beta_0)] \sim \chi^2_1$ , under  $H_0: \beta = 0$ :

Therefore, for a given level of significance  $\alpha$ , we reject  $H_0: \beta = \beta_0$  if

$$2[l(\hat{\beta}) - l(\beta_0)] \geq \chi^2_{1,\alpha}$$

where  $\chi^2_{1,\alpha}$  is the value such that  $P[\chi^2_1 > \chi^2_{1,\alpha}] = \alpha$

Expanding  $l(\beta_0)$  at the MPLE  $\hat{\beta}$ , we get

$$l(\beta_0) \approx l(\hat{\beta}) + \frac{dl(\hat{\beta})}{d\beta}(\beta_0 - \hat{\beta}) + \frac{1}{2!} \frac{d^2l(\hat{\beta})}{d^2\beta}(\beta_0 - \hat{\beta})^2$$

Since MPLE  $\hat{\beta}$  maximizes  $l(\beta)$ , i.e.,

$$U(\hat{\beta}) = \frac{dl(\hat{\beta})}{d\beta} = 0,$$

and

$$\frac{d^2l(\hat{\beta})}{d^2\beta} = -J(\hat{\beta}),$$

so

$$2[l(\hat{\beta}) - l(\beta_0)] \approx J(\hat{\beta})(\beta_0 - \hat{\beta})^2$$

We already derived that

$$(\beta_0 - \hat{\beta}) \sim N(0, J^{-1}(\hat{\beta}))$$

Therefore,

$$2[l(\hat{\beta}) - l(\beta_0)] \approx J(\hat{\beta})(\beta_0 - \hat{\beta})^2$$

$$\left[ \frac{\hat{\beta} - \beta_0}{J^{-1/2}(\hat{\beta})} \right] \sim \chi^2_1 \quad \text{under } H_0: \beta = \beta_0$$

**Cox PH model based on Time-Dependent Covariates**

Sometimes part of the covariates could be time-dependent. For example, the time dependent covariates could be

1. age at failure time t

2. dosage level at failure time t
3. accumulative dosage at failure time t
4. treatment status (o<sup>®</sup> or on) at failure time t
5. or a transformation of the above time-dependent measurements.

Time-dependent covariates for the  $i^{\text{th}}$  individual are

$$x_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))$$

We shall use the general notation  $x_i(t)$  instead of  $x_i$ , even though some of the covariates are time-independent. The PHM is now

$$\lambda(t; x_i(u), u \leq t) = \lambda_0(t) e^{\beta x_i(t)}$$

With time-dependent covariates, the previous partial likelihood argument still works, and the partial likelihood becomes

$$L_P = \prod_{y(i)} \left( \frac{e^{\beta x_i(y(i))}}{\sum_{j \in R(i)} e^{\beta x_j(y(i))}} \right)$$

Example. Suppose

$$x_i(t) = (x_{i1}, x_{i2}(t), x_{i3}(t))$$

$$x_{i1} = \begin{cases} 1 & \text{treatment} \\ 0 & \text{no treatment} \end{cases}$$

$$x_{i2}(t) = \text{the } i^{\text{th}} \text{ individual's age at } t$$

$$x_{i3}(t) = (\text{the } i^{\text{th}} \text{ individual's age at } t)^2$$

$T$  = time from entry to death.

Note that  $x_{i2}(0)$  = baseline age of the  $i^{\text{th}}$  patient. The partial likelihood is

$$L_p = \prod_{y(i)} \left( \frac{e^{\beta_1 x_{i1} + \beta_2 x_{i2}(y(i)) + \beta_3 x_{i3}(y(i))}}{\sum_{j \in R(i)} e^{\beta_1 x_{j1} + \beta_2 x_{j2}(y(i)) + \beta_3 x_{j3}(y(i))}} \right)$$

**Goodness of Fit**

Time-independent  $x$  - material from Miller's book, p168-170

Suppose we want to check on the validity of proportional hazards model. In the case that  $x$  is one-dimensional, an approach of goodness-of-fit is to partition the  $x$ -axis into  $K$  intervals, compute a separate Kaplan-Meier estimate for each interval, then apply the 2-sample goodness of fit procedures. When the time-independent covariate  $x$  is multi-dimensional, we consider the following approach.

Define

$$\Delta_{x_i}(T_i) = e^{\beta x_i} \int_0^{T_i} \lambda_0(u) du$$

Thus, because  $\Delta_{x_i}(T_i)$  is monotonic in  $T_i$ ,

$$P(\Delta_{x_i}(T_i) > t) = P(T_i > \Delta_{x_i}^{-1}(t))$$

$$= \exp(-\Delta_{x_i}(\Delta_{x_i}^{-1}(t))) = e^{-t}$$

Thus, the random variable  $\Delta_{x_i}(T_i)$  follows Exponential ( $\theta = 1$ ) distribution. Further,  $(\Delta_{x_1}(y_1), \delta_1, \dots, (\Delta_{x_n}(y_n), \delta_n))$  form a sample with censoring. Because  $(\Delta_{x_i}(y_i))$  depends on  $\beta$  and  $\lambda_0(t)$ , substitute the corresponding estimates and define

$$\hat{\Delta}_i = \hat{\Delta}_{x_i}(y_i) = e^{\hat{\beta} x_i} \int_0^{y_i} \hat{\lambda}_0(u) du$$

Let  $\hat{S}(t)$  be the Kaplan-Meier estimate based on  $((\hat{\Delta}_1, \delta_1, \dots, (\hat{\Delta}_n, \delta_n))$ . Under the proportional hazards model,  $\log S(t) = -t$  is a linear function of  $t$ . To verify the validity of the proportional hazards model, check if

$\frac{t}{\log \hat{S}(t)} = -1$   
 is approximately satisfied.  
 Time-dependent  $x(t)$

When the covariate  $x(t)$  is time-dependent, the above techniques no longer work for goodness-of-fit. There is a large literature regarding how to construct tests to verify the proportional hazards model assumptions. The so-called 'Martingale residuals' are used as the fundamental statistics for constructing the tests. For continuous survival data, define a 'residual' at  $y(i)$  as

$$r_{(i)} = x_{(i)}(y_{(i)}) - \frac{\sum_{j \in R_{(i)}} x_j(y_{(i)}) \exp(\beta x_j(y_{(i)}))}{\sum_{k \in R_{(i)}} \exp(\beta x_k(y_{(i)}))}$$

$$= x_{(i)}(y_{(i)}) - E[\text{covariate at } y_{(i)} | R_{(i)}]$$

Each residual term has 0 expectation. Thus, after replacing  $\beta$  by  $\hat{\beta}$ , the corresponding residual plot should reflect this specific feature.

**Variables Selection for Cox's Proportional Hazard Models**

The Cox proportional hazard model assumes  $h(t/x) = h_0(t) \exp(x^T \beta)$

where the baseline hazard function  $h_0(t)$  is an unspecified function. To present explicitly the likelihood function of the observed data  $\{(x_i, Z_i, \delta_i): i=1, 2, \dots, n\}$  from Cox's proportional hazards model, more notation is needed. Let  $t_1^0 < \dots < t_j^0$  denote the ordered observed failure times. Let  $(j)$  provide the lable for the item falling at  $t_j^0$  so that the cavariates associate with the N failures are  $x_{(1)}, \dots, x_{(N)}$ . Let  $R_j$  denote the risk set right before the time  $t_j^0: R_j = \{i: Z_i \geq t_j^0\}$ . The likelihood becomes

$$L = \prod_{i=1}^N h_0(Z_{(i)}) \exp(x_{(i)}^T \beta) \prod_{i=1}^n \exp\{-H_0(Z_{(i)}) \exp(x_{(i)}^T \beta)\}$$

where  $H_0(\cdot)$  is the cumulative baseline hazard function. The corresponding penalized log-likelihood function is

$$\prod_{i=1}^N [\log\{h_0(Z_{(i)})\} + x_{(i)}^T \beta] - \sum_{i=1}^n \{H_0(Z_{(i)}) \exp(x_{(i)}^T \beta)\} - n \sum_{j=1}^d p \lambda_n \left( \left| \beta_j \right| \right) \dots (2)$$

Since the baseline hazard and cumulative hazard functions are known and not been parameterized, the panelized log-likelihood function in equ.(2) is not ready optimization yet. Following Breslow's idea, consider the "least informative" nonparametric modeling for  $H_0(\cdot)$ , in which  $H_0(t)$  has a possible jump  $h_j$  at the observed failure time  $t_j^0$ . More precisely, let  $H_0(t) = \sum_{j=1}^N h_j I(t_j^0 \leq t)$ . Then

$$H_0(Z_i) = \sum_{j=1}^N h_j I(i \in R_j) \dots (3)$$

Using logarithm equ.(3), the of penalized likelihood function of equ.(2) becomes

$$\sum_{i=1}^N \{\log\{h_j + x_{(i)}^T \beta\} - \sum_{i=1}^n \{\sum_{j=1}^N h_j I(i \in R_j) \exp(x_{(i)}^T \beta)\} - n \sum_{j=1}^d p \lambda_n \left( \left| \beta_j \right| \right) \dots (4)$$

taking the derivative with respect to  $h_j$  and setting it to be zero, we obtain that

$$\hat{h}_j = \left\{ \sum_{i \in R_j} \exp(x_i^T \beta) \right\}^{-1}$$

substituting  $\hat{h}_j$  into equ. (8.8), we get the penalized likelihood

$$\sum_{i=1}^N [x_{(i)}^T \beta - \log\{\sum_{i \in R_j} \exp(x_i^T \beta)\}] - n \sum_{j=1}^d p \lambda_n \left( \left| \beta_j \right| \right) \stackrel{\text{def}}{=} \ell_c(\beta) - n \sum_{j=1}^d p \lambda_n \left( \left| \beta_j \right| \right) \dots (5)$$

After dropping a constant term "-N". when  $p \lambda(\cdot) \equiv 0$ , equ.(5) is the partial likelihood function suggested by Cox (1975). Thus, the penalized likelihood indeed is the penalized partial likelihood. The penalized likelihood estimate of  $\beta$  is obtained via maximizing equ.(5) with respect to  $\beta$ . For a detailed study on the subject in this direction, refer to Fan et. al. (2005).

**Bayesian Variable selection for Cox's Regression model**

A proportional hazards model is denoted by a hazard function of the form

$$h(t, x) = h_b(t) \exp(x^T \beta)$$

where  $h_b(t)$  denotes the baseline hazard function at time  $t$ ,  $x$  denotes the covariate vector for an arbitrary individual in the population, and  $\beta$  denotes a vector of regression coefficients.

**Prior and Posterior Distribution for  $h_b$**

Let  $G(\alpha, \lambda)$  denote the gamma distribution with shape parameter  $\alpha > 0$  and scale parameter  $\lambda > 0$ , where the density is given by

$$f(x|\alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Consider a discrete gamma process prior for the baseline hazard rate. To denote the discrete gamma process, we first construct a finite partition of the time axis. Let  $0 \leq s_0 < s_1 < \dots < s_m$  denote this finite partition, with  $s_m > t_j$ , for all  $j = 1, \dots, n$ .

$$\text{Let } \xi_i = h_b(s_i) - h_b(s_{i-1})$$

denote the increment in the baseline hazard in the interval  $(s_{i-1}; s_i); i = 1, \dots, M$ . The  $\xi_i$ 's are random variables since the baseline hazard is assumed random. The  $\xi_i$ 's are independent a priori, and have gamma distributions, which are induced by the underlying gamma process. Thus, the  $\xi_i$ 's have independent gamma distributions with shape parameters  $\alpha(s_i) - \alpha(s_{i-1})$  and scale parameter  $\lambda$ . The variance of the gamma process is controlled by choosing  $\lambda$  large or small to reflect vague prior beliefs at various time intervals. Letting  $\tau = (\xi_1, \dots, \xi_m)$ ; the prior density of  $\tau$  is given by

$$P_\tau = \psi(\tau) = \prod_{i=1}^m f(\xi_i)$$

where  $f(\xi_i)$  is a  $G\{\alpha(s_i) - \alpha(s_{i-1}), \lambda\}$  density. The prior parameters of the  $\xi_i$ 's may be chosen to depend on  $m$ . Choices of prior parameters for  $\tau$  can be made in several ways. One may take vague choices of prior parameters for the  $\xi_i$ 's such as  $\alpha(s) = s_i - s_{i-1}$  for  $s_{i-1} \leq s \leq s_i$ , and take  $\lambda$  large. This choice may be suitable if there is little prior information available on the baseline hazard rate.

We assume a priori independence between the baseline hazard rate and the regression coefficients, and thus the joint prior density of  $(\beta^{(m)}, \tau)$  under model  $m$  is given by  $\psi(\beta^{(m)}, \tau|m) = \psi(\beta^{(m)}|m)\psi(\tau)$

We consider a fully parametric multivariate normal prior for  $\beta^{(m)}$ , since the normal prior has proved to be a flexible and useful class of priors for many regression problems, refer to Geisser 1993. Thus let  $N_p(\mu, T)$  denotes the  $p$  dimensional multivariate normal distribution with mean  $\mu$  and precision matrix  $Q$ . Thus, under model  $m$ , we take  $\beta^{(m)} \sim N_{k_m}(\mu^{(m)}, c_0 Q_m)$

where  $c_0$  is a scalar quantifying the degree of prior belief one wishes to attach to  $\mu^{(m)}$ . Under model  $m$ , we have the prior information  $W_0 = (n_0, y_0, X_{0,m}, \delta_0)$  and  $c_0$ , where  $X_{0,m}$  is an  $n_0 \times k_m$  design matrix,  $y_0$  is an  $n_0 \times 1$  vector of prior predictions, and  $\delta_0$  is the corresponding  $n_0 \times 1$  vector of censoring indicators. We take the prior mean of  $\beta^{(m)}$  to be the solution to Cox's partial likelihood equations for  $\beta^{(m)}$  using  $W_0$  as data. Suppose there are  $r$  failures and  $n_0 - r$  right censored values in  $y_0$ . Cox's partial likelihood for  $\beta^{(m)}$  based on  $W_0$  is given by

$$L^*(\beta^{(m)}) = \prod_{i=1}^r \left[ \frac{\exp\{x_{0i}^{(m)'} \beta^{(m)}\}}{\sum_{l \in R(y_{0i})} \exp\{x_{0l}^{(m)'} \beta^{(m)}\}} \right] \quad \dots (6)$$

where  $x_{0i}^{(m)'}$  is the  $i^{th}$  row of  $X_{0,m}$ ,  $(y_{01}, \dots, y_{0r})$  are the ordered failures and  $R(y_{0i})$  is the set of labels attached to the individuals at risk just prior to  $y_{0i}$ . Now we take  $\mu^{(m)}$  to be the solution to

$$\frac{\partial \log\{L^*(\beta^{(m)})\}}{\partial \beta_j^{(m)}} = 0$$

$j = 1, \dots, k_m$ . The matrix  $Q_m$  is taken to be the Fisher information matrix of  $\beta^{(m)}$  based on the partial likelihood in (8.10). Thus

$$Q_m = \left[ \frac{-\partial^2}{\partial \beta_i^{(m)} \partial \beta_j^{(m)}} \log\{L^*(\beta^{(m)})\} \right] \Bigg|_{\beta^{(m)} = \mu^{(m)}}$$

Given the prior prediction  $y_0$ , the prior probability of model  $m$  for the current study based on an update of  $y_0$  via Bayes theorem is given by

$$p(m) = p(m|D_0) = \frac{p(W_0|m)p_0(m)}{\sum_{m \in M} p(W_0|m)p_0(m)} \quad \dots (7)$$

where

$$p(W_0|m) = \iint L(\tau, \beta^{(m)}; W_0) \psi_0(\beta^{(m)}|m) \psi_0(\tau) d\beta^{(m)} d\tau$$

and  $L(\tau, \beta^{(m)}, W_0)$  is the likelihood function of the parameters based on  $D_0$ . The posterior probability of model  $m$  (for the current study) is given by

$$p(m|W) = \frac{p(W|m)p(m)}{\sum_{m \in M} p(W|m)p(m)}$$

where  $p(W|m)$  denotes the marginal distribution of the data given model  $m$ , and  $p(m)$  denotes the prior probability of model  $m$  given by (7). The marginal density  $p(W|m)$  corresponds precisely to the normalizing constant of the joint posterior density of  $(\tau, \beta^{(m)})$ . That is,

$$p(W|m) = \iint L(\tau, \beta^{(m)}; W) \psi(\beta^{(m)}|m) \psi(\tau) d\beta^{(m)} d\tau$$

Finally, we let  $p(\beta, \tau |W)$  denote the posterior distribution of the full model, that is,

$$p(\beta, \tau|W) \propto L(\beta, \tau|W) \psi(\beta|K) \psi(\tau)$$

where  $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ ;  $m = 1, 2, \dots, k$ ;  $k$  is the dimension of the model and is denoted by full model.

**Real data example**

**Cox's proportional hazard model**

Assume the Completed Length of Service (CLS) distribution follows Gamma, the probability density function is

$$f_x(t) = \frac{(\lambda t)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda t}, \lambda > 0, n > 0, t > 0$$

The reliability function is

$$R_x(t) = \frac{1}{\Gamma(n)} \int_{\lambda t}^{\infty} u^{n-1} e^{-u} du, \lambda > 0, n > 0, t \geq 0 \quad \dots (8)$$

which can be recognized as incomplete Gamma function.

$$h_x(t) = \frac{(\lambda t)^{n-1} \lambda e^{-\lambda t}}{\int_{\lambda t}^{\infty} u^{n-1} e^{-u} du}$$

In this case it can be observed that  $h_x(t)$  is an increasing function of  $t$ , when  $n > 1$ . The calculations of  $h_x(t)$  is rather complex as it involves incomplete gamma integrals. However, the expression for  $h_x(t)$  is obtained using computer. First, the least square estimate of the  $\lambda$  and  $n$  are obtained for the stimulated observations.

Using Newton-Raphson iterative procedure, with the statistical analysis and system (SAS) package, the value of  $\hat{\lambda}$  and  $\hat{n}$  for the simulated data set are obtained. Using  $\hat{\lambda}$  and  $\hat{n}$  in equ. (8), and using the estimated values of  $\beta_1, \beta_2, \beta_3$  the propensity of leaving the organization upto the estimated number of years for the different combinations of personal covariates using  $h_x(t)$ , as Gamma distribution for CLS in the case of Cox's proportional hazard model has been estimated. For more detailed discussion the Cox's proportional hazard model using real data example, refer to Elangovan (1998) and Elangovan et. al. (2006).

**Real Data Example**

**Myeloma Study**

A real data example has been taken from Krall et.al. (1975) and Kailash (1992). The former study of Krall et.al. analyze data from study on multiple myeloma in which  $n_0 = 65$  patients with alkylating agent. Of those patients, 48 died during the study and 17 survived. The response variable measured for these data at diagnosis. These are blood urea nitrogen, hemoglobin, platelet count age, white blood cell count, bone fractures, percentage of the plasma cells in bone marrow, proteinuria, and serum calcium. These are typical covariates measured in multiple myeloma studies. Another multiple myeloma study using similar alkylating agents was studied by Kailash (1992), had  $n = 479$  patients with the same set of covariates being measured.

**Table 4** Variable Recorded – Multiple Myeloma Patients

Notation	Variable Name
$t$	Survival time from diagnosis to nearest month +1
$t'$	Maximum observed survival time from diagnosis +1
$A/D$	0 – Alive, 1 – Dead
$X_0$	Constant = 1.0
$X_1$	log BUN at diagnosis
$X_2$	Hemoglobin at diagnosis
$X_3$	Platelets at diagnosis 0 – abnormal, 1 – normal

X <sub>4</sub>	Infection at diagnosis 0 – none, 1 – present
X <sub>5</sub>	Age at diagnosis (complete years)
X <sub>6</sub>	Sex 1 – Male, 2 – Female
X <sub>7</sub>	log WBC at diagnosis
X <sub>8</sub>	Fractures at diagnosis 0 – none, 1 – present
X <sub>9</sub>	log %BM (log % of plasma bone marrow)
X <sub>10</sub>	% Lymphocytes in peripheral blood at diagnosis
X <sub>11</sub>	% Myeloma cells in peripheral blood at diagnosis
X <sub>12</sub>	Proteinuria at diagnosis
X <sub>13</sub>	Bence Jones protein in urine at diagnosis 1 – present, 2 – none
X <sub>14</sub>	Total serum protein at diagnosis
X <sub>15</sub>	Serum globins (gm%) at diagnosis
X <sub>16</sub>	Serum calcium (mgm%) at diagnosis

the equ. Given in (7) with data  $W_0$ . Using the simulation by Gibbs sample, refer to Rubinstein (1981), with the following data values as an input  $c_0 = n_0 / n = 65/339 = 0.27$  and  $(r_1, r_2) = (0.16, 0.19)$ .

**Table 5 Data Set – Multiple Myeloma\***

Case No.	t	t'	A/D	X <sub>0</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>	X <sub>15</sub>	X <sub>16</sub>
1	1.25	40	1	1.0	2.2175	9.4	1.0	0.0	67.0	1.0	3.6628	1.0	1.9542	0.0	0.0	12.0	2.0	11.0	7.0	10.0
2	1.25	51	1	1.0	1.9395	12	1.0	1.0	38.0	1.0	3.9868	1.0	1.9542	0.0	0.0	20.0	1.0	6.0	3.0	18.0
3	2.00	21	1	1.0	1.5185	9.8	0.0	1.0	81.0	1.0	3.8751	1.0	2.0000	0.0	0.0	2.0	1.0	11.0	8.0	15.0
4	2.00	54	1	1.0	1.7482	11.3	0.0	0.0	75.0	1.0	3.8062	1.0	1.2552	13.0	52.0	0.0	2.0	9.0	6.0	12.0
5	2.00	41	1	1.0	1.3010	5.1	1.0	0.0	57.0	1.0	3.7243	1.0	2.0000	0.0	0.0	3.0	1.0	10.0	6.0	9.0
6	3.00	29	1	1.0	1.5441	6.7	1.0	1.0	46.0	2.0	4.4757	0.0	1.9345	3.0	5.0	12.0	2.0	9.0	3.0	10.0
7	5.00	19	1	1.0	2.2355	10.1	1.0	0.0	50.0	2.0	4.9542	1.0	1.6628	8.0	27.0	4.0	1.0	6.0	4.0	9.0
8	5.00	29	1	1.0	1.6812	6.5	1.0	0.0	74.0	1.0	3.7324	0.0	1.7324	1.0	25.0	5.0	2.0	10.0	7.0	9.0
9	6.00	90	1	1.0	1.3617	9	1.0	1.0	77.0	1.0	3.5441	0.0	1.4624	7.0	51.0	1.0	2.0	7.0	5.0	8.0
10	6.00	85	1	1.0	2.1139	10.2	0.0	0.0	70.0	2.0	3.5441	1.0	1.3617	6.0	68.0	1.0	2.0	9.0	7.0	8.0
11	6.00	14	1	1.0	1.1139	9.7	1.0	0.0	60.0	1.0	3.5185	1.0	1.3979	11.0	30.0	0.0	2.0	11.0	7.0	10.0
12	6.00	25	1	1.0	1.4150	10.4	1.0	0.0	67.0	2.0	3.9294	1.0	1.6902	10.0	33.0	0.0	2.0	10.0	6.0	8.0
13	7.00	93	1	1.0	1.9777	9.5	1.0	0.0	48.0	1.0	3.3617	1.0	1.4682	9.0	41.0	5.0	2.0	9.0	3.0	10.0
14	7.00	61	1	1.0	1.0414	5.1	0.0	0.0	61.0	2.0	3.7324	1.0	2.0000	0.0	0.0	1.0	2.0	8.0	4.0	10.0
15	7.00	92	1	1.0	1.1761	11.4	1.0	0.0	53.0	2.0	3.7243	1.0	1.5185	13.0	42.0	1.0	1.0	9.0	6.0	13.0
16	9.00	89	1	1.0	1.7243	8.2	1.0	0.0	55.0	1.0	3.7993	1.0	1.7407	4.0	31.0	0.0	2.0	9.0	6.0	12.0
17	11.00	68	1	1.0	1.1139	14	1.0	0.0	61.0	1.0	3.8808	1.0	1.2788	8.0	49.0	0.0	2.0	8.0	5.0	10.0
18	11.00	15	1	1.0	1.2304	12	1.0	0.0	43.0	1.0	3.7709	1.0	1.1761	7.0	48.0	1.0	1.0	12.0	8.0	9.0
19	11.00	35	1	1.0	1.3010	13.2	1.0	0.0	65.0	1.0	3.7993	1.0	1.8195	4.0	21.0	1.0	2.0	8.0	4.0	10.0
20	11.00	78	1	1.0	1.5682	7.5	1.0	0.0	70.0	1.0	3.8865	0.0	1.6721	5.0	35.0	0.0	2.0	7.0	5.0	12.0
21	11.00	41	1	1.0	1.0792	9.6	1.0	0.0	51.0	2.0	3.5051	1.0	1.9031	5.0	10.0	0.0	2.0	7.0	3.0	9.0
22	13.00	93	1	1.0	0.7782	5.5	0.0	1.0	60.0	2.0	3.5798	1.0	1.3979	12.0	55.0	2.0	2.0	10.0	8.0	10.0
23	14.00	28	1	1.0	1.3979	14.6	1.0	1.0	66.0	1.0	3.7243	1.0	1.2553	7.0	51.0	2.0	1.0	9.0	4.0	10.0
24	15.00	39	1	1.0	1.6021	10.6	1.0	0.0	70.0	1.0	3.6902	1.0	1.4314	16.0	46.0	0.0	2.0	9.0	5.0	11.0
25	16.00	62	1	1.0	1.3424	9	1.0	0.0	48.0	1.0	3.9345	1.0	2.0000	0.0	0.0	0.0	2.0	8.0	4.0	10.0
26	16.00	93	1	1.0	1.3222	8.8	1.0	0.0	62.0	2.0	3.6990	1.0	0.6990	5.0	41.0	17.0	2.0	8.0	4.0	10.0
27	17.00	79	1	1.0	1.2304	10	1.0	0.0	53.0	1.0	3.8808	1.0	1.4472	9.0	50.0	4.0	2.0	10.0	7.0	9.0
28	17.00	60	1	1.0	1.5911	11.2	1.0	1.0	68.0	1.0	3.4314	0.0	1.6128	1.0	46.0	1.0	2.0	9.0	6.0	10.0
29	18.00	49	1	1.0	1.4472	7.5	1.0	0.0	65.0	2.0	3.5682	0.0	0.9031	17.0	46.0	7.0	2.0	6.0	4.0	8.0
30	19.00	44	1	1.0	1.0792	14.4	1.0	0.0	51.0	1.0	3.9191	1.0	2.0000	0.0	0.0	6.0	2.0	13.0	7.0	15.0
31	19.00	52	1	1.0	1.2553	7.5	0.0	0.0	60.0	2.0	3.7924	1.0	1.9294	2.0	10.0	5.0	1.0	6.0	3.0	9.0
32	24.00	39	1	1.0	1.3010	14.6	1.0	0.0	56.0	2.0	3.0899	1.0	0.4771	12.0	52.0	0.0	2.0	4.0	3.0	9.0
33	25.00	56	1	1.0	1.0000	12.4	1.0	0.0	67.0	1.0	3.8195	1.0	1.6435	3.0	23.0	0.0	2.0	7.0	3.0	10.0
34	26.00	34	1	1.0	1.2304	11.2	1.0	0.0	49.0	2.0	3.6021	1.0	2.0000	0.0	0.0	27.0	1.0	6.0	3.0	11.0
35	32.00	48	1	1.0	1.3222	10.6	1.0	0.0	46.0	1.0	3.6990	1.0	1.6335	6.0	32.0	1.0	2.0	12.0	8.0	9.0
36	35.00	66	1	1.0	1.1139	7	0.0	0.0	48.0	1.0	3.6502	1.0	1.1761	16.0	42.0	4.0	1.0	12.0	9.0	10.0
37	37.00	237	1	1.0	1.6021	11	1.0	0.0	63.0	1.0	3.9542	0.0	1.2041	10.0	53.0	7.0	1.0	7.0	4.0	9.0
38	41.00	58	1	1.0	1.0000	10.2	1.0	0.0	69.0	1.0	3.4771	1.0	1.4771	5.0	30.0	6.0	1.0	10.0	6.0	10.0
39	41.00	73	1	1.0	1.1461	5	1.0	0.0	70.0	2.0	3.5185	1.0	1.3424	18.0	53.0	0.0	2.0	7.0	4.0	9.0
40	51.00	61	1	1.0	1.5682	7.7	0.0	0.0	74.0	1.0	3.4150	1.0	1.0414	10.0	50.0	4.0	1.0	12.0	8.0	13.0
41	52.00	57	1	1.0	1.0000	10.1	1.0	0.0	60.0	2.0	3.8573	1.0	1.6532	1.0	27.0	4.0	1.0	10.0	7.0	10.0
42	54.00	94	1	1.0	1.2553	9	1.0	0.0	49.0	1.0	3.7243	1.0	1.6990	8.0	37.0	2.0	1.0	8.0	3.0	10.0
43	58.00	74	1	1.0	1.2041	12.1	1.0	0.0	42.0	2.0	3.6990	1.0	1.5798	9.0	35.0	22.0	1.0	7.0	4.0	10.0
44	66.00	74	1	1.0	1.4472	6.6	1.0	0.0	59.0	1.0	3.7853	1.0	1.8195	4.0	22.0	0.0	2.0	5.0	3.0	9.0
45	67.00	97	1	1.0	1.3222	12.8	1.0	0.0	52.0	1.0	3.6435	1.0	1.0414	24.0	54.0	1.0	1.0	11.0	7.0	10.0
46	88.00	90	1	1.0	1.7161	10.6	1.0	0.0	47.0	2.0	3.5560	0.0	1.7559	22.0	17.0	21.0	1.0	6.0	3.0	9.0
47	89.00	121	1	1.0	1.3222	14	1.0	1.0	63.0	1.0	3.6532	1.0	1.6232	7.0	38.0	1.0	1.0	9.0	7.0	9.0
48	92.00	11	1	1.0	1.4314	11	1.0	0.0	58.0	2.0	4.0755	1.0	1.4150	12.0	36.0	4.0	1.0	9.0	6.0	11.0
49	4.00	4	0	1.0	1.9542	10	1.0	0.0	59.0	1.0	4.0453	0.0	0.7782	21.0	51.0	12.0	1.0	7.0	3.0	10.0
50	4.00	4	0	1.0	1.9243	10	1.0	1.0	49.0	2.0	3.9590	0.0	1.6232	21.0	24.0	0.0	2.0	17.0	12.0	13.0
51	7.00	7	0	1.0	1.1139	12.4	1.0	0.0	48.0	2.0	3.7993	1.0	1.8573	1.0	11.0	0.0	2.0	6.0	3.0	10.0
52	7.00	7	0	1.0	1.5315	10.2	1.0	1.0	81.0	1.0	3.5911	0.0	1.8808	1.0	16.0	0.0	2.0	12.0	10.0	11.0
53	8.00	8	0	1.0	1.0792	9.9	1.0	0.0	57.0	2.0	3.8325	1.0	1.6532	1.0	40.0	0.0	2.0	8.0	4.0	8.0
54	12.00	12	0	1.0	1.1461	11.8	1.0	0.0	46.0	2.0	3.6435	0.0	1.1461	7.0	47.0	0.0	2.0	7.0	3.0	7.0
55	11.00	11	0	1.0	1.6128	14	1.0	0.0	60.0	1.0	3.7324	1.0	1.8451	0.0	0.0	3.0	1.0	6.0	2.0	9.0
56	12.00	12	0	1.0	1.3979	8.8	1.0	0.0	66.0	2.0	3.8388	1.0	1.3617	8.0	50.0	0.0	2.0	9.0	5.0	9.0
57	13.00	13	0	1.0	1.6628	4.9	0.0	0.0	71.0	2.0	3.6435	0.0	1.7924	2.0	21.0	0.0	2.0	10.0	11.0	9.0
58	16.00	16	0	1.0	1.1461	13	1.0	0.0	55.0	1.0	3.8573	0.0	0.9031	13.0	53.0	0.0	2.0	9.0	5.0	9.0
59	19.00	19	0	1.0	1.3222	13	1.0	0.0	59.0	2.0	3.7709	1.0	2.0000	0.0	0.0	1.0	2.0	8.0	4.0	10.0
60	19.00	19	0	1.0	1.3222	10.8	1.0	0.0	69.0	2.0	3.8808	1.0	1.5185	0.0	0.0	0.0	2.0	10.0	7.0	10.0
61	28.00	28	0	1.0	1.2304	7.3	1.0	1.0	82.0	2.0	3.7482	1.0	1.6721	7.0	40.0	0.0	2.0	7.0	4.0	9.0
62	41.00	41	0	1.0	1.7559	12.8	1.0	0.0	72.0	1.0	3.7243	1.0	1.4472	2.0	56.0	1.0	1.0	7.0	3.0	9.0
63	53.00	53	0	1.0	1.1139	12	1.0	0.0												

**Table 6** The models with the largest posterior probabilities for various values of  $c_0$ 

$c_0$	model	p(m)	Model	p(m D)
0.19	( $x_1, x_2, x_3, x_7$ )	0.10	( $x_1, x_2, x_3, x_4, x_7, x_8$ )	0.286
0.50	( $x_1, x_2, x_3, x_7$ )	0.10	( $x_1, x_2, x_3, x_4, x_7, x_8$ )	0.394
1.00	( $x_1, x_2, x_3, x_7$ )	0.10	( $x_1, x_2, x_3, x_4, x_7, x_8$ )	0.284

The real data example shown in previous section using proportional hazard model based on three covariates only. If more number of covariates are used, in such a case computation will be so difficult and it is observed that most of the covariate combinations, the propensative to leave increase with time. It is not so in the case of computing prior and posterior probability using multiple myeloma study. The prior and posterior probability can be calculated only moderate variable selecting problem involving 15 – 20 covariates. For more than 20 covariates and for large regression problem, the method suggested by George and McCulloch (1993) can be adopted for the proportional hazard model. As is well known that in the case of Cox's proportional hazard model and Bayesian variable selection model based on the real data example, it is observed that both the model are perhaps equally suitable. Based on the Myeloma data example the simulation results showed that the posterior models probabilities were not sensitive to the increasing hazard rate assumption and this is not to be true in general. Once the figures or information's are made available, the application of the models would be really helpful for future decision. Hence it shows that Bayesian variable selection for Cox's regression model performs effectively. Finally the proposal methodology (in both proportional hazard and Bayesian variable selection) can be used as a tool for screening suitable subsets of models which could be consider for future analysis.

\*\*\*\*\*

## Reference

1. Cox, D.R. (1970). The analysis of Binary Data. London : Methuen.
2. Cox, D.R. (1972). Regression model and life tables (with discussion). *Journal of the Royal Statistical Society*, B 34, 187-220.
3. Cox, D.R. (1975). Partial likelihood, *Biometrika*, 62, 269-296.
4. George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, 88, 884-889.
5. Kalish, L. A. (1992). Phase III multiple myeloma: evaluation of combination chemotherapy in previously untreated patients. Technical Report # 726E, Department of Biostatistics, Dana-Farber Cancer Institute, Boston, MA.
6. Krall, J.M., Uthoff, V.A. and Harley, J.B. (1975). A Step-Up Procedure for Selecting Variables Associated with Survival, *Biometrics*, 31(1), 49-57.
7. Gilks, W. R., and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, 41, 337-48.
8. Elangovan, R., N. Vijayasankar and R. Sathiyamoorthy. (2006). Application of Cox's Regression Model in the Analysis of Wastage of Manpower Data. *Annamalai University Journal of Science*, Vol.43, 173-186.
9. Elangovan, R. (1998). A Study on some Stochastic Model for Manpower Planning. Ph.D. Thesis submitted to Annamalai University, India.
10. Fan, J. and Li, R. (2005). Contemporary Multivariate Analysis and Design of Experiments. World Scientific, New Jersey, pp. 315-336.